EFFECTIVENESS OF DATA BREACH LEGISLATION, 2005-2012

A Thesis
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Master of Public Policy
in Public Policy

By

Caroline E. Dwyer, B.A.

Washington, DC
April 22, 2014
EFFECTIVENESS OF DATA BREACH LEGISLATION, 2005-2012

Caroline E. Dwyer, B.A.

Thesis Advisor: Micah Jensen, P.h.D.

ABSTRACT

This thesis examined the effectiveness of mandatory data breach reporting legislation and the penalties said laws sometimes impose for failing to report a data breach. I examined data breach data gathered for 2005-2012 and the relevant state legislation using state-year fixed effects regression. My analysis suggested that the existence of a mandatory data breach reporting law is associated with higher numbers of reported breaches. Also, reporting laws which impose a penalty for not reporting may be more effective than laws which don't, but the effect of the penalty depends on the amount of the penalty itself. Penalties which are too low may not adequately compel organizations which have been attacked to report the data breach.
The research and writing of this thesis is dedicated to everyone who helped along the way. Special thanks to Mom, Dad, and Kevin - without your patience and support, this thesis never would have happened.

Many thanks,
Caroline E. Dwyer
# CONTENTS

INTRODUCTION .................................................................................................................................................. 6
BACKGROUND & LITERATURE REVIEW ...................................................................................................... 9
CONCEPTUAL MODELS AND THEORETICS .................................................................................................. 12
DATA & METHODS ......................................................................................................................................... 15
RESULTS ......................................................................................................................................................... 19
DISCUSSION ................................................................................................................................................... 24
CONCLUSION .................................................................................................................................................. 28
APPENDIX ...................................................................................................................................................... 30
REFERENCES .................................................................................................................................................. 33
INTRODUCTION

It's the phone call everyone dreads making. The one to your credit card company, asking why your MasterCard didn't work at Trader Joe's. You listen to awful elevator music for thirty minutes, before you're granted the privilege of talking to some grumpy bank employee. When you finally are allowed to speak to a human being, she asks if you really did buy five thousand plastic pink flamingo decorations on November 15th, 2013. Of course you say, "No, I didn't do that." That's when you find out someone stole your identity. That's when you find out what a data breach is; that it doesn’t matter how careful you are, or that your bank is careful too. At any rate, you definitely need to find out if you can be reimbursed for those five thousand flamingos.

For those unfamiliar, a data breach is an event like the one Target recently suffered. It is defined as "an unauthorized acquisition of computerized data that compromises the security, confidentiality, or integrity of PI (personal information) maintained by the Entity," where an "entity" is any organization; this definition comes from California's data breach reporting legislation (Perkins Coie, 2013). I used California's definition here because California has one of the most stringent data breach reporting laws currently on the books. In layman's terms, a data breach occurs when someone illegally gains access to people's private information where it is stored on a computer. Data breaches are common - even expected by businesses - anywhere that information related to bank accounts, credit cards, health insurance and healthcare, social security, or state and national security might be located (DeVost, 2013).

Data breaches, therefore, are important in that they mark a significant loss of privacy and security. Data breaches can lead to a wide variety of problems for both institutions attacked and individuals whose data is stolen: the most well-known problems are identity theft, fraud,
financial theft, intellectual property theft, and in the case of government data breaches, loss of military advantages in areas such as defense and weapons technology. It is the purpose of this paper to test the effects of data breach reporting legislation on actual reporting rates. Data breach reporting legislation aims to force entities to report data breaches when they occur – reporting breaches provides customers and clients, as well as analysts, with additional information so that breaches can be learned from and either prevented or dealt with more efficiently in the future. This legislation also encourages entities to do everything in their power to prevent breaches, because making breaches known can be damaging to these entities. Specifically, in this paper I test three theories: (1) the implementation of mandatory reporting regulation increases the number of breaches reported, (2) reporting laws that impose a penalty for non-compliance are more likely to increase the number of breaches reported, and (3) the amount of the penalty associated influences the number of breaches reported.

These hypotheses are important to examine because the penalties in each state for not reporting may not serve as an adequate incentive to report data breaches. Alaska, for instance, only charges a maximum of $50,000 in total fines per incident, with the possibility for the $50,000 fine capped annually regardless of number of incidents. Some states have higher penalties, and others have an even lesser penalty - and some states have no penalties whatsoever. Given that even in many states that have penalties, the amounts are small (for the business world), these laws might not be adequate enough to convince entities to report breaches. In light of this, it is important to study the effectiveness of this legislation.

Another factor to consider is the lack of federal-level legislation. Even the Office of Management and Budget's Memorandum M-07-16, *Safeguarding Against and Responding to the*
Breached Personally Identifiable Information, which requires government agencies (and any affiliated contracting companies) to report breaches, is not official legislation passed in Congress - merely an advisory memorandum from OMB. It may be that such guidance is better than none, but it is possible or even likely that properly constructed reporting laws are more effective tools to fight data breaches.

Some researchers have asserted that reporting legislation is necessary in order to ascertain quantity and severity of data breaches. Widup (2010) and Collins et al. (2011) suggest such a conclusion is accurate. Moreover, there are definite gaps and unknowns in the literature and available evidence: much of the data is limited or unavailable for public analysis, and what is available often lacks certain key variables. Further, it is impossible to know if all data breaches have been accounted for in such data, why certain data was attacked, or why the attack was of a certain type.

I propose that, while the implementation of mandatory breach reporting legislation may be good, it must have “teeth” to be effective. A piece of legislation without an adequate incentive for compliance accomplishes nothing. The legislation therefore should have compelling and significant penalties if ignored. Legislation that involves no such mechanisms may be ignored, and might result in still more insecure systems and security vulnerabilities that could expose thousands of customers’ data. For this reason, I chose to study how differences in the penalty aspects of data breach reporting laws might affect the effectiveness of the law in compelling institutions to report when they have been attacked.
BACKGROUND & LITERATURE REVIEW

Understanding current literature regarding data breaches and data breach legislation is key to understanding the hypotheses this thesis will test. But since data breaches and relevant legislation are such a comparatively recent development, existing literature is somewhat limited, and data breaches and data breach reporting legislation are not yet researched in adequate depth. It is the purpose of this section to examine the literature available on data breaches and data breach reporting legislation.

Theory and Literature:

In this paper, I test whether mandatory data breach reporting legislation is effective, and whether the penalties involved have a significant impact. This is important for two main reasons: firstly, there is no overarching federal legislation mandating data breach reporting; secondly, in many of the states that do have mandatory reporting legislation, there exist only small penalties for breaking the law, if there are penalties at all. As data breaches represent a significant threat, it is very important that they be carefully reported, tracked, and measured; yet there is no existing comprehensive federal legislation that requires data breaches to be reported. Instead, the legislation is piecemeal, with federal laws requiring only the healthcare sector (Stevens, 2012, Health, 2013), the Veterans Administration (Stevens, 2012, Panangala, 2006), government agencies, and government contracting industries to report data breaches (Stevens, 2012, Johnson, 2007, Faulkner, 2007). Additional legislation, known as the Gramm-Leach-Bliley Act (GLBA), requires financial institutions (banks, investment companies, insurers, and brokerages) to report; however, it was passed in 1999, and was not written with data breaches in mind (Faulkner, 2007).
It is a further problem that data breaches invariably go unreported, since the existing state-level legislation does not provide a strong enough incentive to report breaches (Winn, 2009). Winn’s 2009 article "Are 'Better' Security Breach Notification Laws Possible?" argues that there are many pitfalls in existing security breach notification laws (these are the same as data breach reporting laws); her theories call into question the effectiveness of these laws on companies, and whether they really do provide adequate incentives to most companies to report breaches. Winn later explains California's trailblazing reporting legislation; this is highly important since California's law served as a template for later states' reporting laws, and the flaws that existed in the California template became pervasive across states. Such a flaw would be exemption when encrypted data is involved (Winn, 2009). She further attempts to test whether or not legislation could be altered to increase effectiveness; her conclusion advocates for federal legislation to close the gaps and vulnerabilities caused by state laws.

Flora Garcia's 2007 article argues contrary to Winn's point, and posits that (when her article was published in 2007) the nation was not yet ready for federal legislation on data breach reporting. Garcia concluded that more time was needed to learn about data breaches and different legal approaches to mandating breach reporting. She does note that federal legislation will eventually become a wise and necessary step, but she qualifies this statement with the idea that data breach reporting legislation must be explored and better understood before the federal government can effectively and efficiently deal with this issue (Garcia, 2007). Now that several years have passed and a number of high profile data breaches have raised public awareness of this issue, research on the efficacy of state and federal legislation regarding data breach reporting is needed.
Suzanne Widup's 2010 article represents another major work regarding data breaches. Widup's article discusses frequency of data breaches, records disclosed, breach vectors, criminal use, data types, relationships and costs; the article also contains sections which discuss organizational sectors and geographic location of breaches. It is this last section which is most useful to this thesis. According to Widup, the business sector represents the largest number of incidents - but these incidents are on the decline as of 2009. The education, government, and medical sectors are all comparable. Verizon's 2013 Data Breach Investigations Report confirms that the business sector has the highest incidence of data breaches (Verizon, 2013). Widup's data is further supported by Collins, Sainato, and Khey's 2011 article. This article seconds Widup's findings in many ways, and examines the same time period's data. While it focuses on the healthcare and education sectors, it also notes the other sectors affected by data breaches, and includes all sectors in its analysis of data breaches between 2005 and 2010.

Limitations of Previous Literature:

Faulkner, Garcia, and Winn's articles suggest a common weakness of the existing literature - they were written between four and six years prior to this study and many states have implemented legislation since those articles were written. However, as noted earlier, much of this legislation is a result of states copying California's legislation, so it is not known if the effect of the new reporting legislation has been maximized. Also, Widup's article, while a seminal work, only examines data breaches between 2005 and 2009, and since its initial publication, it has only been updated to include data from 2010. This leaves several years of unexamined data, and given how rapidly the technology industry tends to advance, study of those years might yield critical new information.
As the Verizon Annual Data Breach Investigation Report of 2013 notes, it is impossible to know everything about data breaches - many go unreported due to lack of legislation and fear; many more go unreported because only the attacker is aware a breach has taken place. This too highlights a weakness in all available research on data breaches. Not much is known about the unreported breaches. This is a further reason to study the configuration of reporting laws, so that future data breaches will not remain hidden and unstudied.

**Theory:**

This thesis will tests the theories that: (1) the implementation of mandatory reporting regulation increases the number of breaches reported, (2) reporting laws that impose a penalty for non-compliance are more likely to increase the number of breaches reported, and (3) the amount of the penalty associated influences the number of breaches reported. Further methodology will be addressed in the Data and Methods section of this proposal.

**CONCEPTUAL MODELS AND THEORETICS:**

Data breaches are the result of malicious attacks on vulnerable systems, or systems that are assumed to be secure but have previously- undiscovered vulnerabilities, yet the number and severity of these attacks is difficult to measure. This is due in part to the fear organizations feel when they discover they have been attacked - vulnerability and insecurity is not something that makes customers or clients trust an organization, and so it might sometimes try to conceal a breach. This is because the data exposed is often sensitive data - personal identification information or sensitive company or government information, usually stored digitally.
Legislation that mandates data breach reporting forces businesses to reveal any breaches, which may motivate businesses to take steps to prevent future breaches.

Businesses generally report more breaches than other types of organizations. The private sector is leading in information security, not the government. According to Professor DeVost, this is widely known in the industry (DeVost, 2013). The business sector had the foresight to realize that happy customers feel aware of the possible threats to their information security, and feel protected from it. Therefore the business sector instituted an industry standard of reporting data breaches. The government subtly endorsed this industry standard when it passed the Federal Trade Commission Act, which "prohibits unfair and deceptive practices in and affecting commerce (Stevens, 2012, Federal, 2011)." It is further supported by the creation of the Payment Card Industry Data Security Standards (PCI-DSS, 2006), which sets the standard for information security of companies that deal with bank cards (Stevens, 2012, Payment, 2010). The business sector also seems to have realized that in reporting these breaches, it would have to take extra steps to make sure its customers felt protected. This also was the purpose of the PCI-DSS and FTC Act. This combination of reporting and additional security measures was likely designed to make customers happier. Happier customers meant fewer companies try to hide data breaches, because of the risk involved in hiding the breaches. This ideally would mean that more businesses report data breaches, since it benefits their customers and therefore their businesses as well.

Why might businesses not want to report breaches, and why might data on breach reporting be misleading? It’s plausible that companies and organizations may prefer to hide breaches, fearing that alerting customers to a weakness will damage the company. This is a result
of what happened to CitiBank, when it disclosed a data breach it felt was fairly well-contained and countered – when CitiBank disclosed the breach, its stocks plummeted and customers abandoned the company out of fear; other companies learned from CitiBank that they might be better off hiding data breaches from customers (DeVost, 2013). But after CitiBank's revelation, customers wanted to know about such vulnerabilities; they wanted legislation. As Widup notes, only four states lacked data breach reporting laws as of 2010: they are New Mexico, Alabama, Kentucky, and South Dakota. The remaining 46 states, along with the District of Columbia, Puerto Rico, and the Virgin Islands have data breach reporting laws which vary in stringency (Stevens, 2012). Since state-level legislation forcing the reporting of breaches was then implemented in 46 states, it seems likely that more breaches were exposed and that people were more aware of the vulnerabilities companies face regarding data breaches. Widup notes that the two states which, as of 2010, had long-standing and stringent data breach reporting laws also represented the largest percentages of data breaches across the nation. This impacts how one must interpret statistics on data breaches - it seems unlikely that they are experiencing more data breaches - merely that more data breaches are being reported in those states. Because of this growing movement towards legislation, coupled with customers' desire to be in-the-know, companies have had to become better at compensating and protecting their customers.

If there is no mandatory reporting legislation, or if the penalties for not complying with the legislation are not adequate, the law may be ignored and customers may be left vulnerable, but it is necessary to study the effectiveness of such reporting legislation, and particularly the influence of penalties for non-compliance in order to identify features that may impact the effectiveness of any future federal-level legislation. In this study, I expected to find evidence that
states with reporting legislation report more attacks than states that do not have such legislation - thereby indicating that there is some effect of such legislation. If so, the benefit would be that attacks are reported, instead of being hidden and unknown to the public, which could be hazardous to individuals whose information has been compromised. I also expected to find that, for relevant legislation to have any impact on the situation, it must have a penalty, and one that is compelling, not merely a nuisance. I believe that showing how lesser penalties are less effective will be important so that law makers can set penalties that encourage obedience of the laws. However, the amount of the penalty also matters because if the amount of the penalty is too small organizations may ignore it.

**DATA & METHODS**

My primary hypothesis is that legislation has a significant impact on reported cases of data breach. If this is true, then more incidents should be expected to be reported in states with mandatory reporting laws. My secondary hypothesis is that laws that include a penalty cause an increase in the number of incidents reported. If this is true, then states with no penalties will not report lower or higher numbers of breaches than states with penalties, on average. My third hypothesis is that the amount of the penalty matters. If this is true, then states with higher penalties will report more breaches than states with small or no penalties.

To state these hypotheses formally:

1: H1: States with data breach reporting legislation do not report more breaches than states without such laws.
H0: States with data breach reporting legislation report more breaches than states without such laws.

2:  H2: States with data breach reporting legislation that includes a penalty report more breaches than states whose reporting laws do not have penalties.

     H0: States with data breach reporting legislation that includes a penalty do not report more breaches than states whose reporting laws do not have penalties.

3:  H3: States with (higher) maximum penalties for failing to obey data breach reporting legislation do not report more breaches than states with lower penalties.

     H0: States with (higher) maximum penalties for failing to obey data breach reporting legislation report more breaches than states with lower penalties.

Description of the Data:

To test these three hypotheses, data from the Privacy Rights Clearinghouse was used (Privacy Rights Clearinghouse, 2013). This data encompasses data breaches between 2005 and 2012; there are approximately 2,000 observations. The unit of analysis is state-year. The data set includes variables for the date the breach was made public, the year the breach occurred, the name of the victim organization, the sector the organization is a part of, the kind of cyber attack used, the city and state of the victim organization, the total records held by the organization, the total records breached in the attack, what was stolen, how appropriate parties were notified of the breach, and how many attached files were involved. There are 407 observations in this sample that are useable. The Privacy Rights Clearinghouse compiled its data primarily from
datalossDB.org, but also used data from various news articles, Google Alerts, SANS (a security e-newsletter), the California Attorney General’s Office, the International Association of Privacy Professionals Daily Dashboard, and Phiprivacy.net. While the Privacy Rights Clearinghouse freely admits its dataset does not include all data breaches, it does note that it serves as a useful indication of the situation surrounding data breaches, data breach reporting, and relevant legislation.

The second dataset I used is compiled from two charts; one put out by the Mintz Levin law firm, and the other put out by Perkins Coie, a company that provides legal counsel for companies (Perkins Coie, 2013; Mintz Levin, 2013). The relevant sections of the Mintz Levin chart details whether or not laws exist in each of the 50 states, and describes the penalties involved. It measures them in dollars per affected state resident, and lists the total maximum penalty per incident or per year (as appropriate). If there is no specific penalty listed, it lists whether there is truly no penalty, if civil penalties might apply, and whether state fair trade laws apply. The relevant sections of the Perkins Coie chart lists the dates that mandatory data breach legislation took effect in each state. Both charts also list the four states (Alabama, Kentucky, New Mexico, and South Dakota) that do not have laws related to data breach notification as of July 1, 2013. The sample size for these charts is 50 - one entry for each state. There was no sampling strategy used by either of these data collectors - they collected information regarding the legal documents on data breach notification in each of the 50 states.

The third dataset is from the Bureau of Economic Analysis. It is a state and year based list of GDP. I used this data to calculate a second dependent variable. I divided breaches reported by the state’s GDP in order to account for differences in the size of each state’s economy.
Additional variables created and used in the regressions include variables to indicate the presence of mandatory reporting regulations without penalties, regulations with penalties, the existence of regulation, and the amount of penalties.

The dependent variables were the existence of regulation, regulation without penalties, regulation with penalties, the existence of a maximum penalty, the amount of the penalty, and the amount of the penalty squared.

**Regression Model:**

I performed a state-year fixed effects analysis, using panel data that shows reported data breaches in all 50 states between 2005 and 2012. The formal equation for a state-year fixed effects model is shown below:

\[ y_{it} = \beta_1 x_{it} + \alpha_i + e_t + u_{it} \]

Where \( y \) is the number of breaches per year (t) in a given state (i). The \( \beta \) represents the beta coefficients for the independent variables. The \( \alpha \) variables represents the state fixed effects. The \( e \) variable represents the year fixed effects. The \( u \) variable represents error from both state and year for the regression.

In a series of separate regressions, I regressed the number of breaches in a given state in a given year and the number of breaches divided by state GDP against the following variables: number of breaches, existence of any mandatory reporting regulations, existence of a maximum penalty for non-compliance, amount of the maximum penalty and amount of the maximum penalty squared.

**Limits of this Data:**
The datasets on data breaches are very difficult to obtain, and in most cases are not publicly available. All datasets on this topic are incomplete, because there is no way to know about ALL data breaches - some are not reported, others are never even discovered. The dataset I used has only a very few limited number of variables. This means many possible confounding factors go unaccounted for - such as total financial loss from a breach, if the legislation (should it exist in a given state) is enforced, how many files were targeted in the breach, whether the entity in question had protective measures in place, how strong those measures were, security measures put in place at the attacked businesses, size of the companies attacked (although that can be googled), and motive for the attack. Due to difficulty finding available (non-classified) data, these variables cannot be controlled for, and that allows for the possibility that the regression estimates are biased. The results should be interpreted with caution.

RESULTS

To examine the hypotheses put forth earlier in this paper, I estimated eight regression models, paired off into four overarching sets with alternative forms of the dependent variable. The first set of models considered whether the existence of any regulation in a state reduces the number of breaches reported in that state. The second set of models examined questions about the significance of penalties when regulation is present. The third set explored the differences between a lack of regulation and regulations that include penalties; it also explored the differences between penalty amounts. The fourth and final set sought to evaluate the impact of different penalty amounts. Estimates for all models are presented in Table 2, which can be found appended to the end of this document.
Models 1A & 1B:

This model is two bi-variate regressions following the equations listed below. Its results determine the effect of mandatory data breach reporting laws of any kind, on the number of breaches actually reported in a state. The regression equation for model 1A is:

\[
\text{breaches} = \hat{\beta}_0 + \hat{\beta}_1 \text{ any regulation}
\]

The equation's dependent variable is number of total breaches. The independent variable is whether or not a state has any regulation in effect in a given year between 2005 and 2012. The coefficient for the any regulation variable is statistically significant, as seen in column 1A of Table 2; it is also highly significant with a p-value of less than 0.01. This tells us that the presence of regulation is associated with roughly 3.9 more reported breaches in a given year than in states without legislation. I then ran the following, subsequent regression (model 1B):

\[
\text{breaches per GDP} = \hat{\beta}_0 + \hat{\beta}_1 \text{ any regulation}
\]

This equation's dependent variable is number of total breaches divided by state GDP. The independent variable is again whether or not a state has any regulation in effect in a given year between 2005 and 2012. The coefficient for the any regulation variable is significant, as seen in column 1B of Table 2; it is also highly significant with a p-value of less than 0.01. This tells us that 1.74 more breaches per $100,000 GDP are reported in states that have mandatory data breach reporting regulations, than in states that do not have such regulations. This tells us that states with any mandatory data breach reporting legislation do report more breaches, even
adjusting for the size of the state's economy. It is important to note that the results from models 1A and 1B are strikingly similar.

Models 2A & 2B:

These models add additional controls and involve F-tests for equality of coefficients. These models examine the effect of non-compliance penalties as part of data breach reporting laws. The regression equation for model 2A is:

$$\text{breaches} = \hat{\beta}_0 + \hat{\beta}_1 \text{regulation with no penalty} + \hat{\beta}_2 \text{regulation with penalty}$$

This equation's dependent variable is the total number of breaches. The independent variables represent regulations with no penalties, and regulations with penalties, respectively. The coefficient for the regulation with no penalty variable is 3.724, as seen in column 2A of Table 2. It is highly significant with a p-value of less than 0.01. The coefficient for the regulation with a penalty variable is 4.038, and is very significant with a p-value of less than 0.01. An F-test for the equality of coefficients finds that coefficients are not significantly different from each other (F = 0.06, d.f. = 405, p-value= 0.8141). This suggests that the effect of reporting laws may not differ for the inclusion of a penalty for non reporting, but I explore this unexpected result further below.

But I must examine whether or not this is true when GDP is taken into account. I use the below regression for model 2B:

$$\text{breaches per GDP} = \hat{\beta}_0 + \hat{\beta}_1 \text{regulation with no penalty} + \hat{\beta}_2 \text{regulation with penalty}$$
The coefficient for the regulation with no penalty variable is now considerably smaller at $2.05 \times 10^{-5}$, as seen in column 2B of Table 2. It is highly significant with a p-value of less than 0.00. The coefficient for the regulation with a penalty variable is also much smaller, at $1.64 \times 10^{-5}$, as seen in column 4 of Table 1. It is highly significant with a p-value of less than 0.01. These coefficients are more surprising still as they suggest that, after adjusting for GDP, laws with a penalty have a weaker, positive effect than laws without a penalty, on number of reported breaches but again these coefficients are not statistically different from one another ($F = 0.40$, d.f. = 405, p-value = 0.5309).

**Models 3A & 3B:**

These models seek to examine the effect of the amount of the penalty. The regression equation for model 3A is:

$$
\text{breaches} = \hat{\beta}_0 + \hat{\beta}_1 \text{regulation with no penalty} + \hat{\beta}_2 \text{amount of maximum penalty}
$$

The coefficients, standard errors, and p-values (both of which are highly significant) can be found in column 3A of Table 2. From these results I estimate that compared to states without reporting requirements, states which have regulation without penalty, report 3.349 more breaches. There is also a predicted 0.00000384 rise in number of reported breaches for each dollar increase in the penalty amount; in other words, for every $100,000, the data predicts a 0.384 increase in reported breaches.

The equation for model 3B is:

$$
\text{breaches per GDP} = \hat{\beta}_0 + \hat{\beta}_1 \text{regulation with no penalty} + \hat{\beta}_2 \text{amount of maximum penalty}
$$
The coefficients here are again difficult to interpret as they are in units that cannot be compared to one another. The coefficients, standard errors, and p-values (both of which are highly significant) can be found in column 3B of Table 2. It should be noted that these coefficients are much smaller than in the first regression of this model.

**Models 4A & 4B:**

These models examine the influence of the penalty amount. Here I squared the amount of the maximum penalty to allow for a non-linear relationship with the dependent variables, and treated regulation with no penalty as an indicator variable, and amount of maximum penalty and amount of maximum penalty squared as continuous variables. The regression equation for model 4A is:

\[ \text{breaches} = \hat{\beta}_0 + \hat{\beta}_1 \text{ regulation with no penalty } + \hat{\beta}_2 \text{ amount of maximum penalty } + \hat{\beta}_3 \text{ amount of maximum penalty}^2 \]

Coefficients, p-values, and standard errors can all be found in column 4A of Table 2. From these results I estimate that compared to states without reporting requirements, states which have regulation without penalty, there is a positive 3.349 increase in total number of reported breaches. For states with penalties, the amount of the penalty does itself influence the effectiveness of legislation at a statistically significant level (F-test of joint significance: F=49.0, d.f. = 50, p-value < 0.001). When I investigated these results more deeply, I discovered that there is a “sweet spot” for penalties. This can be seen in Figure 1, appended to the end of this document. Figure 1 presents the estimated average marginal effects for the range of maximum penalty amounts. It can be seen in the figure that the effect of the non-reporting penalty is only
statistically significant between approximately $600,000 and $1,600,000. The confidence
intervals outside that range overlap zero and so the penalty does not appear to be statistically
significant when it is either too low or too high.

I then ran the regression with breach per GDP as the dependent variable, to see if the
results were substantively consistent. The below equation was used to generate model 4B:

\[
\text{breaches per GDP} = \hat{\beta}_0 + \hat{\beta}_1 \text{ regulation with no penalty} + \hat{\beta}_2 \text{ amount of maximum penalty} + \\
\hat{\beta}_3 \text{ amount of maximum penalty}^2
\]

Coefficients, p-values, and standard errors can all be found in column 4B of Table 2. From these
results I estimate that compared to states without reporting requirements, states which have
regulation without penalty, there is a 1.9 increase in total number of reported breaches per
$100,000 of state GDP. In states whose reporting laws include a non-reporting penalty, the
marginal effect of all penalties below $1,550,000 were as shown in Figure 2, appended to the end
of this document. Again, the penalty amount was a significant predictor of breaches per GDP (F-
test of joint significance: F=22.07, d.f. = 50, p-value < 0.001).

**DISCUSSION**

As described in the results section, each of the four models have unique implications.
This section's purpose is to give additional detail and analysis of those implications. The overall
meaning of the four models shows that the mere existence of legislation has the highest positive
impact on breach reporting. The existence of a penalty does not appear significant until its effect
is specified in a non-linear fashion; then the amount of a penalty also has an impact.
I began with my most basic overarching curiosity: by attempting to ascertain whether or not mandatory data breach reporting legislation is effective. All eight regressions support the idea that reporting laws are associated with higher numbers of reported breaches, and models 1A and 1B show this most directly, suggesting that mandatory reporting legislation may have a large, positive, meaningful effect on number of breaches reported. This is good – this is what I expected to see. But I must note two caveats. Firstly, it is possible that the direction of causation could be reversed, and that instead states with a higher likelihood of breaches were simply more likely to pass reporting laws. Further research could be done on this in the future to determine the direction of causality. Secondly, these results might on the surface seem like regulations increase the number of breaches that occur; I believe that this is not the case, and that the legislation is merely exposing breaches that would have happened anyway, but would have gone unreported in the absence of legislation. My data and analysis cannot, however, prove or disprove this conclusion.

I next sought to determine the importance of penalties as part of mandatory reporting laws. I wanted to know if, for states that have laws, the presence of a penalty for not reporting a breach makes a difference in the number of breaches reported. Models 2A and 2B ideally would have answered this question. However their results do not support the conclusion that the inclusion of a penalty in such a law has any meaningful effect. Further, the results from models 2A and 2B were contradictory in that they indicated different magnitudes for the influence of the inclusion of a penalty. Additionally, their differences were not statistically significant. But I suspected that the specification of these models did not properly reflect the true nature of the relationships, and so I created models 3 and 4.
Models 3 and 4 sought to determine whether, for states that have laws and penalties, the amount of the penalty itself makes a difference in reported breaches. I suspected that the effect of a penalty for non-reporting varied by the amount of the penalty. I postulated that companies might run a cost-benefit analysis and discover that the cost of the penalty was less than the cost of obeying the mandatory reporting legislation, and might therefore simply refuse to report because the penalty was so small it made it more economical to ignore the law than to obey it. Models 3A and 3B suggest this may indeed be so. In fact, estimates in model 3A suggest (as noted earlier in the Results section) that there is an ideal range for penalties between $600,000 and $1,600,000. Any laws with a penalty up to $600,000 would produce approximately the same increase in reported breaches as do the reporting laws that do not have a penalty. This is particularly interesting since $600,000 is more than double the maximum penalty for reporting laws in states that have such a feature. This data implies several possibilities: that low penalties are not enough of a disincentive, that the average state penalty may be too low, and that the effect of a penalty may not be linear. This last implication prompted me to generate models 4A and 4B.

Models 4A and 4B allow the penalty amount to vary in a non-linear fashion by including controls for both the penalty amount and the penalty amount squared. These models suggest that low penalty amounts do not seem to affect breach reporting rates. But very high penalty amounts seem to suggest similar ineffectiveness. I believe this result is somewhat misleading, and that instead extremely high penalty amounts are so effective that they incentivize organizations to reduce their risk of breach as much as possible, thereby lessening the number of breaches that take place and could be reported. This possible ineffectiveness of low penalties, and theoretically
extreme effectiveness of high penalties brings us to the data of models 4A and 4B. The data here suggests that there is a “sweet spot” for penalty amounts, between $600,000 and $1,600,000, where there is a positive impact on number of breaches reported.

**Weaknesses & Limits of Research:**

Unfortunately the models I examined have a several potential weaknesses. The data itself is limited, and sometimes awkward to quantify. The data is limited in that it is incomplete, because better data sets were classified or unavailable; the nature of the data is also a weakness, in that data breaches often go unreported and therefore cannot be accounted for in even the most complete data sets on the topic. Also, some states, for instance, do not set simple maximum penalties, but rather escalate the penalties as time continues to pass; others set penalties per affected individual - this makes it difficult to truly accurately account for each state's penalties. The collinearity and contradictory results seen in model 2 is also a weakness - there may exist some better way of examining the importance of penalties that is more effective - but I've yet to find it. Model 4 is particularly difficult to pin down, as well - squaring or logging the effect of the amount of the penalty presented difficulties, as the meaning of the results I obtained were difficult to interpret.

This paper also has certain limits - conclusions and generalizations that go beyond the scope of my hypotheses should not be made based on this research. Most importantly, any conclusions that deal with specific economic sectors - business vs government vs medical vs education, etc - should be avoided. This paper lumps them all together, and does not delve into the differences between them. This is because the legislation that applies to government and
medical sectors is different than state legislation, and also came into effect midway through the
time period this paper examines. It would be far more complex of a study to undertake, were I to
attempt to answer any questions along those lines.

CONCLUSION

This paper set out to examine the importance and impact of mandatory data breach
legislation and penalties associated with that legislation, on the number of actual data breaches
reported.

Specifically this paper looked at the relative impacts of mandatory reporting legislation,
mandatory reporting legislation that included penalties, and the amount of those penalties on data
breach reporting in all 50 states. This paper presents data and analysis that suggests both the
existence of a regulation and the inclusion of a non-reporting penalty are significant, but that
penalty’s effect depends on its amount. The existence of regulation may have a greater effect
than a low penalty, but the impact of penalties may increase as the penalties increase. Further,
there appears to be a maximum penalty amount after which effectiveness drops off.

Future research possibilities include replicating the results of this study, preferably with
more robust data. Lending credit to this study would help build a stronger argument for federal-
level legislation on data breach reporting. Such legislation would, ideally, encourage entities
with sensitive information to be more careful and to increase and keep up to date their cyber
security measures. I feel that this would go a long way towards securing digitized sensitive
information in the US. Additional research could be done on the effects of legislation on
different sectors of the economy, and what effect existing federal-level legislation has on data
breaches and on data breach reporting. Further research could also be done regarding the impact of penalty amounts - what amounts are too little incentive, how much is too much, and if there is an optimum penalty amount. This is particularly important for existing state-level legislation – the amendment of penalty amounts so that they fall within the maximally effective range might help make these mandatory reporting laws more effective.
## APPENDIX
### TABLE 1: Variables and Relevant Descriptive Statistics

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>DESCRIPTION</th>
<th>RANGE</th>
<th>OBSERVATIONS</th>
<th>CENTER</th>
<th>STANDARD DEVIATION</th>
<th>SHAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total breaches</td>
<td>Total number of breaches</td>
<td>0 - 119</td>
<td>408</td>
<td>Mean: 8.72</td>
<td>12.11</td>
<td>Peaks in 2006 and 2012, with valleys in 2005 &amp; 2009</td>
</tr>
<tr>
<td>Breach per $1 of GDP</td>
<td>Total number of breaches per $1 of state GDP</td>
<td>3.58e-6 - 0.0003</td>
<td>408</td>
<td>Mean: 0.00</td>
<td>0.00</td>
<td>Peaks in 2006 and 2011, with valleys in 2005 &amp; 2009</td>
</tr>
<tr>
<td>Any regulation</td>
<td>Whether or not mandatory data breach reporting regulations are in effect in any given state in any given year</td>
<td>0 - 1</td>
<td>407</td>
<td>Mode: 1</td>
<td>0.44</td>
<td>Skewed - more legislation appears as time passes</td>
</tr>
<tr>
<td>Regulation but no penalty</td>
<td>Indicates a state that has mandatory data breach reporting regulations in effect, but that the regulations do not carry a penalty</td>
<td>0 - 1</td>
<td>408</td>
<td>Mode: 0</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Regulation with penalty</td>
<td>Indicates a state that has mandatory data breach reporting regulations in effect, and that the regulations carry a penalty</td>
<td>0 - 1</td>
<td>408</td>
<td>Mode: 0</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>Maximum penalty</td>
<td>Indicates the existence of a maximum penalty in a given state</td>
<td>0 - 1</td>
<td>408</td>
<td>Mode: 0</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Amount of maximum penalty</td>
<td>Indicates the maximum monetary penalty regulations carry in a given state in a given year</td>
<td>0 - 2,960,000</td>
<td>408</td>
<td>Mean: 97780.88</td>
<td>401666.20</td>
<td></td>
</tr>
<tr>
<td>VARIABLES</td>
<td>MODEL 1:</td>
<td></td>
<td>MODEL 2:</td>
<td></td>
<td>MODEL 3:</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------------------</td>
<td>-------------------------</td>
<td>------------------</td>
<td>-------------------------</td>
<td>------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td></td>
<td>(1A)</td>
<td>(1B)</td>
<td>(2A)</td>
<td>(2B)</td>
<td>(3A)</td>
<td>(3B)</td>
</tr>
<tr>
<td>Total</td>
<td>3.724***</td>
<td>2.05***</td>
<td>4.038***</td>
<td>0.164***</td>
<td>3.894***</td>
<td>1.74***</td>
</tr>
<tr>
<td>breaches</td>
<td>(1.074)</td>
<td>(0.546)</td>
<td>(0.918)</td>
<td>(0.401)</td>
<td>(0.752)</td>
<td>(0.335)</td>
</tr>
<tr>
<td>GDP</td>
<td>Total</td>
<td>3.84e-06***</td>
<td>0***</td>
<td>1.20e-05*</td>
<td>4.47e-11**</td>
<td></td>
</tr>
<tr>
<td>Breaches/</td>
<td>(9.99e-07)</td>
<td>(0)</td>
<td>(6.96e-06)</td>
<td>(2.24e-11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$100,000</td>
<td>GDP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Amount max penalty</td>
<td>3.84e-06***</td>
<td>0***</td>
<td>1.20e-05*</td>
<td>4.47e-11**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9.99e-07)</td>
<td>(0)</td>
<td>(6.96e-06)</td>
<td>(2.24e-11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Amount max penalty$^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.17e-12)</td>
<td>(-1.31e-17*)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.34e-12)</td>
<td>(7.52e-18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>5.886***</td>
<td>2.63e-05***</td>
<td>5.879***</td>
<td>2.56e-05***</td>
<td>7.169***</td>
<td>3.13e-05***</td>
</tr>
<tr>
<td></td>
<td>(0.551)</td>
<td>(2.45e-06)</td>
<td>(0.543)</td>
<td>(2.55e-06)</td>
<td>(0.385)</td>
<td>(1.86e-06)</td>
</tr>
<tr>
<td>Observations</td>
<td>407</td>
<td>407</td>
<td>408</td>
<td>408</td>
<td>408</td>
<td>408</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.040</td>
<td>0.058</td>
<td>0.041</td>
<td>0.061</td>
<td>0.017</td>
<td>0.028</td>
</tr>
<tr>
<td>Number of stateid</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>51</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
FIGURE 1:

Average Marginal Effects of Max Penalty Amount with 95% CIs

Max Penalty Amount

FIGURE 2:

Average Marginal Effects of Max Penalty Amount with 95% CIs

Max Penalty Amount
REFERENCES


Winn, Jane K. "Are 'Better' Security Breach Notification Laws Possible?" (June 8, 2009):